AD_____

Award Number: DAMD17-96-1-6084

TITLE: Metastic Regulation of Differential Splicing of CD44

PRINCIPAL INVESTIGATOR: Susan Berget, Ph.D.

CONTRACTING ORGANIZATION: Baylor College of Medicine
Houston, Texas 77030

REPORT DATE: August 2000

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	August 2000	Annual (15 Jul 99 - 14 Jul 00)

4. TITLE AND SUBTITLE
Metastic Regulation of Differential Splicing of CD44

5. FUNDING NUMBERS
DAMD17-96-1-6084

6. AUTHOR(S)
Susan Berget, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Baylor College of Medicine
Houston, Texas 77030

E-MAIL:
sberget@bcm.tmc.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
This report contains colored photos

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT ***(Maximum 200 Words)***

Alternative mRNA splicing is a major step in gene expression used to produce different forms of a protein for specific cellular purposes. The process is normally carefully regulated to limit production of different mRNA isoforms to only appropriate tissues. Cancer, however, alters splicing regulation and causes the appearance of forms of mRNAs and their protein products not normally present in a cell. The cell surface receptor CD4, is a protein that undergoes extensive alternative processing and whose processing alters in both breast cancer tumors and their metastases. We have been investigating this splicing of several alternative exons within the CD44 gene to understand how the splicing of these exons is regulated and how that splicing alters during tumorigenesis and metastasis. We have observed sequences within two alternatively recognized exons that are the binding sites for two important RNA-binding proteins – human Tra 2 and the y-box protein YB1. Both proteins bind these exon sequences. This binding may be significant for exon recognition because raising the *in vivo* concentration of either protein enhances CD44 alternative splicing. More importantly the levels of both proteins incrrease in mammary tumors and their metastases making them candidate factors responsible for altering CD44 splicing during tumorigenesis.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Breast Cancer	42
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

SB Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

SB In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and use of Laboratory Animals of the Institute of Laboratory Resources, national Research Council (NIH Publication No. 86-23, Revised 1985).

____ For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

____ In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

____ In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

____ In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

Susan Bryant 8-03-00

PI - Signature Date

Table of Contents

(5) INTRODUCTION

Alternative splicing is a mechanism used to increase the number of mRNAs produced from a single gene. The cell adhesion molecule CD44 undergoes extensive alternative splicing (8, 10, 21). Ten CD44 alternative exons, all in the same translational reading frame, reside within a single block in the middle of the gene (Fig. 1). Interest in CD44 splicing peaked when it was observed that inclusion of one alternative exon was causative for metastasis in a rat model of colon cancer (12). Since that time a number of studies have appeared suggesting that alternative CD44 splicing was a useful prognostic marker for breast cancer detection and prognosis in humans. A number of laboratories including our own has observed that CD44 alternative splicing increases in cancer cells (3, 8, 26). There are simultaneous alterations in expression of certain alternative splicing factors, including the well-studied serine-arginine-rich (SR) proteins (10, 26). SR proteins are known to bind to splicing enhancer sequences located within exons and thereby stimulate the binding of constitutive splicing factors such as U2 snRNP-associated factor (U2AF) and U1 snRNPs (1, 20, 23).

One important SR protein is the transformer 2 protein (Tra 2) (14, 17, 30). Discovered in *Drosophila melanogaster*, Tra 2 regulates alternative splicing in the *D. melanogaster* doublesex gene by binding to AC-rich enhancer sequences (18) within the alternative exon in conjunction with another SR protein, Tra. Two genes very similar to *D. Melanogaster* Tra 2, Tra 2α and Tra 2β, exist in humans (6, 28). Both have the capacity to bind to purine-rich exon enhancer sequences although a natural target gene for either human isoform of Tra 2 has been unknown. Last year we reported that human

Tra 2α enhances exon inclusion of CD44 variable exons 4 and 5 in HeLa cells. In contrast, other SR proteins, including SRp55, ASF/SF2, and SC35 had no effect on the splicing of these two exons. These results suggested that human Tra 2 proteins could play a role in alternative processing of CD44 RNA. This observation became more interesting when we observed that the expression of murine Tra 2 protein increased during mammary tumorigenesis.

During the last year we have extended these studies to investigate the mechanism whereby Tra 2 proteins affect CD44 splicing. We observed that addition of recombinant Tra 2 to *in vitro* splicing extracts made from HeLa cells caused the increased binding of another protein to CD44 exon v4 sequences. This protein was identified as the y-box single stranded nucleic acid binding protein YB1 (27, 31). Both Tra 2 and YB1 appeared to bind to CD44 exon 4 sequences at the same time suggesting the existence of a complex containing both proteins. Transfection of HeLa cells with an expression plasmid producing human YB1 pronouncedly activated the inclusion of CD44 variable exons v4 and v5. Mutant versions of YB1 containing the RNA-binding domain but lacking accessory domains could not activate CD44 splicing. The effect was only observed when the reporter plasmid contained introns ruling out a role for YB1 in the stability of CD44 mRNA. Both Tra 2 and YB1 are known to bind to AC-rich sequences. To identify which sequences within CD44 exon v4 bound Tra 2 and YB1 we created versions of our CD44 mini-gene in which the three major AC-rich regions of the exon were mutated (Fig. 1). Mutation of any of the three sequences decreased the binding of YB1. Mutation of the third AC-rich sequence dramatically altered both *in vitro* spliceosome assembly and *in vivo* inclusion of exon v4. We conclude that YB1 and Tra 2 participate in the

activation of CD44 exon v4 recognition. Like Tra 2, the levels of YB1 increase during mammary tumorigenesis and are high in human breast cancer (2), suggesting that alterations in both YB1 and Tra 2 may be responsible for the changes in CD44 splicing observed in breast cancer.

(6) BODY

Experimental Methods, Assumptions, and Procedures

Plasmids and transfections. The reporter CD44 mini-gene used throughout this study (diagrammed in Figure 8) was created by inserting sequences from the human CD44 gene into the first exon of an engineered β-globin *in vivo* expression plasmid, Dup 33 (7) obtained from R. Kole, University of North Carolina, Chapel Hill, NC. The CD44 sequence spanned a sequence from 792 nucleotides before variable exon 4 to 515 nucleotides after variable exon 5 and included all of the natural exon between exons 4 and 5. The YB1 protein expression plasmids were obtained from T. Cooper, Department of Pathology, Baylor College of Medicine.

All transfections used HeLa cells and were done in the presence of lipofectAMINE™ (Gibco/BRL) according to the manufacturer's instructions. Total cell RNA was isolated 48 hours post transfection using TRIzol™ (Gibco/BRL) following manufacturer's instructions. Splicing patterns were determined by RT/PCR analysis using 5' end-radiolabeled primers specific for β-globin sequences (5' AGACACCATGCATGGTGCACC and 3' CCTGATCAGCGAGCTCTAG). These primers amplified no RNA from untransfected HeLa cells. Amplification conditions were 1 min. at 94°C, 1 min. at 58°C and 1.5 min at 72°C for 25 cycles. Product DNA was

denatured and displayed on a 6% urea gel. RNA products were quantified in the Phosphoimager. Plasmid pBR322 digested with Hpa II markers is used in all displayed gels for product size determination. Identified amplification products resulting from the inclusion of one or two variable CD44 exons were sequenced to verify identity.

Plasmids used to generate RNAs for *in vitro* work were created by cloning appropriate regions of the CD44 exon v4 downstream of the T7 or SP6 promoters. For *in vitro* assembly or UV cross-linking RNAs were made that included 38 nucleotides of intron 8, intact exon v4 and intact intron 9. Transcripts were synthesized *in vitro* by cleavage of the DNA within exon v4 at the ava II site located within exon v4 at 49 nucleotides downstream of the 3' splice site or cleavage downstream of exon 4v at an Sfa NI site located 56 nucleotides downstream of the 5' splice site. The former transcript included ACE 1 and ACE 2 but not ACE 3; the latter contained all three. The ACE 1, ACE 2, and ACE 3 mutations were created by PCR mutagenesis. All constructs were sequenced to verify CD44 sequences.

***In vitro* assays.**

In vitro splicing, spliceosome assembly, and UV cross-linking assays have been described previously (4). For experiments involving recombinant proteins, purified YB1 and Tra 2 alpha or beta were gifts of T. Cooper (Baylor College of Medicine) or W. Mattox (MD Anderson), respectively. Antibodies specific for YB1 were provided by T. Cooper (Baylor College of Medicine). Immunoprecipitations were as described previously (4).

Results and Discussion

The Binding of Tra 2 to CD44 Variable exon 4 activated the binding of the human y-box protein YB1 to CD44 sequences.

The sequence of CD44 variable exon 4 is shown in Figure 1. Within the exon are three AC-rich (denoted in red in Figure 1) sequences that resemble the AC-rich exon enhancer from the *D. melanogaster* doublesex gene known to bind to Tra 2 (14, 17, 18)(Figure 2). The CD44 consensus sequence derived from the repeats in exons v4 and v5 is CAACCA. The Tra 2 binding sequence from the doublesex gene is CAAUCAACA, which is similar but not identical, suggesting that proteins other than Tra 2 may be involved in recognition of CD44 exon v4.

To examine proteins that bind CD44 exon v4 we investigated the proteins that can be UV cross-linked to exon 4 sequences in an *in vitro* splicing extract. We noticed that a prominent protein of 50 kDa cross-linked to an RNA containing the first 49 nucleotides of exon v4 using either an S100 extract (Figure 3, right) or a nuclear extract (Figure 4, left). This protein cross-linked considerably better when we added recombinant Tra 2 to the S100 extract, suggesting an interaction with Tra 2 (Figure 3, right, compare lanes 1 and 2) The stimulation appeared to be relatively specific for Tra 2 because other splicing proteins could not cause this effect, including a preparation (32) that contains many of the abundant well known SR proteins (the SR pellet, prepared as described in **).

A 50 kDA protein known to bind AC-rich sequences in the y-box single stranded nucleic acid binding protein YB1 (27, 28). This protein was isolated as the protein that binds to a AC-rich exon enhancer selected as a sequence causing increased exon

inclusion in an *in vivo* selection assay (T. Cooper, manuscript in preparation). The sequence to which this protein binds is **CACCA**GCTACGCGTC**CACCA.** We noticed that the YB1 binding sequence contained two copies of the sequence CACCA, similar to the AC-rich consensus sequence we derived for CD44 exon v4. To see if the 50 kDA protein we were observing in our UV cross-linking studies was indeed YB1, we immunoprecipitated cross-linking reactions with a YB1-specific antibody (Figure 4, left). The 50 kDa cross-linked band was effectively immunoprecipitated with this antibody indicating that it was indeed YB1.

To confirm an interaction between Tra 2 and YB1 we asked if recombinant Tra 2 would activate the binding of recombinant YB1 to exon 4 sequences in the absence of extract (Figure 3, left). Even small amounts of Tra 2α strongly activated the binding of YB1. The binding of both full length YB1 and a proteolytic fragment of YB1 present in the preparation of YB1 used for the experiment was activated. Considerably more cross-linking of YB1 than Tra 2 was observed in this experiment (note that cross-linked Tra 2 runs at a migration just slightly faster than the YB1 proteolytic fragment and can really only be seen at the highest level of Tra 2 employed). Because we do not know the specific activity of either the YB1 or Tra 2 preparations it is difficult to assess the relative activity of the two proteins but it is obvious that even small amounts of Tra 2 activated YB1 binding either in the presence or absence of extract proteins.

RNA binding experiments must be confirmed with competition studies because of the high degree of non-specific RNA binding exhibited by most of these proteins. Figures 5 and 6 show competition of UV cross-linking of YB1 to exon 4 sequences using nuclear extract, a substrate consisting of the first half of exon 4 and containing the first

two AC-rich elements (hereafter referred to as ACE 1 and ACE 2), and a competitor RNA derived from the exon enhancer element selected by the Cooper laboratory and binding YB1 (referred to in figures as the AC-rich SELEX winner). In Figure 5 all of the proteins cross-linking to the exon in an *in vitro* nuclear splicing extract are shown. Increasing amounts of competitor RNA containing the two CACCA motifs effectively competed the cross-linking of YB1. Figure 6 shows the same experiment after immunoprecipitation of cross-linked proteins with the YB1-specific antibodies. In the experiment in Figure 6 we also included the cross-linking pattern of a substrate in which the wild type ACE 1 sequence contained introduced point mutations (see below). The cross-linking of YB1 to this substrate was reduced compared to wild type but could also be competed. We also show the ability of the AC-rich competitor to compete cross-linking of YB1 to itself (Figure 6, lanes 12-16). Lower concentrations of competitor were effective in competing cross-linking of YB1 to the CD44 exon 4 sequences than to the AC-rich SELEX winner sequence, presumably reflecting the difference in overall sequence of CD44 exon 4 and the AC-rich SELEX sequence and the participation of additional proteins in recognition of the former in nuclear extracts. It should be noted that when the Cooper laboratory isolated YB1 as a protein binding the AC-rich SELEX sequence, it was the only nuclear protein that they could detect demonstrating efficient binding to this sequence, suggesting that other proteins do not play a role in recognition of this sequence. Thus, we would anticipate a more complicated pattern of protein binding to the CD44 exon sequences than to the AC-rich SELEX winner and thereby a slightly different response to competition.

Both Tra 2 and YB1 bind to exon 4 at the same time

The increased cross-linking of YB1 in the presence of Tra 2 suggested that both could bind to the RNA at the same time. To directly test this hypothesis, we used gel-shift experiments. Radiolabeled RNA templates containing the portion of exon 4 containing the ACE 1 and ACE 2 elements were incubated with mixtures of recombinant proteins and the formed complexes were visualized on native gels. Complexes appear as areas of retarded radioactivity. Addition of either YB1 or Tra 2β caused complex formation. Addition of YB1 and Tra 2α or Tra 2β caused the appearance of complexes of higher gel migration than the complexes formed with either YB1 or Tra 2β alone, suggesting the creation of complexes containing both YB1 and Tra 2.

YB1 activates exon 4 inclusion *in vivo*

The *in vitro* experiments described above suggest that YB1 is involved in exon 4 recognition. To monitor the effect of YB1 on CD44 splicing, we co-transfected expression plasmids coding for wild type or a mutant version of YB1 into HeLa cells along with a reporter construct containing CD44 variable exon 4 and 5 placed within a human beta-globin gene (23). This same reporter responds to co-transfection with either Tra 2α or Tra 2β with increased inclusion of both exons v4 and v5 (Figure 8, reprinted here from last year's report for comparison). A similar experiment with YB1 indicated that YB1 was even more effective than either human Tra 2 isoform for stimulating exon v4 and v5 inclusion (Figure 9). Inclusion rose from 28% to over 80%. Little RNA was produced that included either v4 or v5 alone without the other variable CD44 exon. Instead, both exons were included. A mutant form of YB1 truncated approximately half

way through the protein (diagrammed in Figure 9) was unable to cause full activation of exon inclusion. This mutant protein contains the cold shock domain (CSD) which is thought to be the nucleic acid binding domain of the protein (27, 31) but is lacking other domains. In its ability to perform wild type function, the mutant YB1 behaves like truncated SR proteins which also fail to function when their SR domains have been deleted.

Neither YB1 or Tra 2 was able to affect the inclusion of a heterologous weak exon inserted into the same mini-gene backbone (Figure10). This exon lacked notable AC-rich sequences and would not be expected to respond to either protein. YB1 was also unable to influence RNA levels when we transfected a CD44 cDNA construct in which exons v4 and v5 had already been spliced. For this experiment we created a mini-gene from the product of splicing of the mini-gene shown in Figure 9. This mini-gene was driven from the same promoter and terminated with the same poly(A) site as the mini-gene in Figure 9; it differed only in the absence of introns. The fact that addition of YB1 had no effect on levels of RNA produced from this mini-gene indicated that the effect in Figure 9 was unlikely to have been caused by effects on mRNA stability. We cannot rule out the possibility of an indirect effect of YB1 on CD44 splicing caused by the transcriptional up-regulation of an unknown splicing protein by the increased levels of YB1. Coupled with the *in vitro* demonstrations of the ability of YB1 to bind to CD44 exon sequences and the ability of Tra 2 to stimulate this binding, the most straightforward interpretation of the data in Figure 9 is that YB1 directly participates in the splicing of CD44 variable exons v4 and v5.

The exon 4 ACE 1 and ACE 2 elements are required for maximal splicing and binding of YB1.

If YB1 is important for exon 4 recognition, it would be predicted that the multiple AC-rich sequences within exon 4 would be required for activity. To test this hypothesis three mutants were prepared in which one of the three AC-rich elements within exon v4 were mutated. The introduced mutations are diagrammed in Figure 11. When these mutants were transfected into HeLa cells in the context of the mini-gene described in Figure 9, two of the three mutations affected observed levels of inclusion of exons v4 and v5. The ACE 3 mutation had the strongest phenotype. It depressed the percentage of product RNA arising from inclusion of both exons v4 and v5 from ___ %to __% of the total RNA. Instead a major new product appeared of length appropriate to be a product RNA in which one but not both of the CD44 exons was included. This product was sequenced and found to be RNA that included only variable exon v5. Thus the mutation had no effect on inclusion of v5 but caused depressed inclusion of exon v4.

The ACE 2 mutation had little effect suggesting that it is not a major sequence regulating exon 4 inclusion. The ACE 1 sequence had an unusual effect. When mutated, exon v4 and v5 inclusion was increased suggesting that this sequence represents a modest exon silencer. A silencer with a different sequence has been experimentally detected in the beginning of exon 5 as well (16). We are currently constructing double and triple mutations to see if mutation of the ACE 1 and ACE 2 elements has a stronger phenotype in the background of an ACE 3 mutation.

We also examined the phenotypes of the ACE 1 and 3 mutations *in vitro*. Figure 12 displays the ability of the ACE 1 and 3 mutants to direct *in vitro* spliceosome

assembly using a substrate RNA containing intact exon 4 with its surrounding splice sites but no other sequences. Such substrates have the ability to assembly the first ATP-dependent spliceosome complex, complex A. As can be seen from the figure, wild type RNA forms complex A, as does the ACE 1 mutant. The ACE 3 mutant is depressed for complex A formation in agreement with the *in vivo* depression of production of spliced RNA containing exon 4. This result suggests that at least a portion of the phenotype of an ACE 3 mutation *in vivo* is due to a reduced ability to form the first spliceosome complex.

We have also begun to analyze the binding of YB1 to the three ACE 1 mutants. Figure 13 displays a UV cross-linking experiment using the wild type and ACE 1 mutant RNAs. Using nuclear extract, the ACE 1 mutant displays lowered ability to be cross-linked to YB1 as compared to wild type RNA. Interestingly, this difference in activity is not observed using cytoplasmic extracts. We suspect that the latter phenotype may be due to the absence of Tra 2 in cytoplasmic extracts and are currently preparing recombinant Tra 2 to add to cytoplasmic extracts to see if the presence of Tra 2 causes the appearance of phenotype for the ACE 1 mutant.

YB1 is induced during mammary tumorigenesis

The inclusion of exon v4 is induced upon mammary tumorigenesis and in a number of human cancers, including breast cancer and metastasis. If YB1 is involved in this increased splicing of the variable CD44 exons during cancer one might predict alterations in the level of expression of YB1 during tumorigenesis. As reported last year we are using a mouse model of mammary development and tumorigenesis that permits

analysis of normal tissues, developmentally regulated normal tissues, preneoplasias and tumors and metastases thereof. We examined the relative level of YB1 mRNA in these different stages by RT/PCR analysis of total cell RNA (Figure 14). No YB1 mRNA was detectable in mature virgin gland or in gland from pregnant animals. Considerable RNA was detected in lactating gland, suggesting induction of YB1 during development. We were unable to detect YB1 in a number of tested preneoplasias. In contrast multiple tumors or their metastases demonstrated considerable YB1 induction. Our results with murine tissue is similar to recent studies indicating high levels of YB1 in human breast cancer tissue (2). Thus, similar to Tra 2, YB1 levels increase in tumor cells, consistent with the possibility that they play a role in induction of CD44 alternative exon splicing during tumorigenesis.

(7) KEY RESEARCH ACCOMPLISHMENTS

- ✓ Determination that recombinant Tra 2α activates the binding of the y-box protein YB1 to CD44 exon sequences
- ✓ Determination that human YB1 influences CD44 alternative splicing *in vivo*
- ✓ Determination that at least two of the AC-rich sequences within exon v4 are important for *in vivo* splicing of exon 4, *in vitro* assembly of exon 4 and YB1 binding.
- ✓ Determination that YB1 is induced during mouse mammary tumorigenesis.

(8) REPORTABLE OUTCOMES

Manuscripts:

We are in the process of assembling the work on Tra 2 and YB1 into several manuscripts. Because of the relationship we discovered between YB1 and Tra 2 activation of CD44 variable splicing we decided to wait and produce multiple manuscripts rather than publish pieces of the story historically.

(9) CONCLUSIONS

Our results indicate that alternative splicing of CD44 is a complicated process involving multiple factors that recognize important sequences within the specific exons and boost splicing. Here we report the interaction of two very different types of RNA binding proteins – human Tra 2 and the y-box protein YB1 in the recognition of regulatory sequences within the alternative exon v4 of CD44. The correlation of binding, in vivo activation, and mutation analysis of AC-rich sequences that bind these two proteins is strong evidence that these proteins play a role in CD44 alternative splicing. Furthermore, both proteins are induced in mammary tumorigenesis, indicating that they play an important role in the up-regulation of CD44 alternative splicing seen during this process.

(10) REFERNCES

1. **Adams, M. D., D. Z. Rudner, and D. C. Rio.** 1996. Biochemistry and regulation of pre-mRNA splicing. **8:**331-339.

2. **Bargou, R. C., K. Jurchott, C. Wagener, S. Bergmann, S. Metzner, K. Bommert, M. Y. Mapara, K. -J. Winzer, M. Dietel, B. Dorken, and H. -D. Royer.** 1997. Nuclear localization and increased levels of transcription factor YB-1 in primary human breast cancers are associated with intrinsic MDR1 gene expression. Nature Medicine 3:447-478.

3. **Cannistra, S., G. Abu-Jawdeh, J. Niloff, T. Strobel, L. Swanson, J. Andersen, and C. Ottensmeier.** 1995. CD44 variant expression is a common feature of epithelial ovarian cancer: lack of association with standard prognostic factors. J. Clin. Oncol. **13:**1912-1921.

4. **Carlo, T., D. A. Sterner, and S. M. Berget**. 1996. An intron enhancer containing a G-rich repeat facilitates inclusion of a vertebrate micro-exon. RNA, **2:**342-353.

5. **Coulter L. R., M. A. Landree, and T. A. Cooper.** 1997. Identification of a new class of exonic splicing enhancers by in vivo selection. Mol Cell Biol **17:**2143-2150.

6. **Dauwalder, B., F. Amaya-Manzanares, and W. Mattox.** 1996. A human homolog of the Drosophila sex determination factor transformer-2 has conserved splicing regulatory functions. Proc. Natl. Acad. Sci. USA **93:**9004-9009.

7. **Dominski, Z., and R. Kole.** 1992. Cooperation of pre-RNA sequence elements in splice site selection. Mol. Cell Biol. **12:**2108-2114.

8. **East, J. A., and I. R. Hart.** 1993. CD44 and its role in tumour progression and metastasis. Eur. J. Cancer **29A(14):**1921-1922.

9. **Fox, S. B., J. Fawcett, D. G. Jackson, I. Collins, K. C. Gatter, A. L. Harris, A. Gearing, and D. L. Simmons.** 1994. Normal human tissues, in addition to some tumors, express multiple different CD44 isoforms. Cancer Res. **54:**4539-4546.

10. **Ghigna, C., M. Moroni, C. Porta, S. Riva, and G. Biamanti.** 1998. Altered expression of heterogeneous nuclear ribonucleoproteins and SR factors in human colon adenocarcinomas. Cancer Res. **58:**5818-5824.

11. **Gunthert, U.** 1993. CD44: a multitude of isoforms with diverse functions. Curr. Top. Microbiol. Immunol. **184:**47-63.

12. **Gunthert, U., M. Hofmann, M. Rudy, S. Reber, M. Zoller, I. Haubmann, S. Matzku, A. Wenzel, H. Ponta, and P. Herrlich.** 1991. A new variant of glycoprotein CD44 confers metastatic potential to rat carcinoma cells. Cell **65:**13-24.

13. **Haynes, B. F., H.-K. Liao, and K. L. Patton.** 1991. The transmembrane hyaluronate receptor CD44: Multiple functions, Multiple forms. Cancer cells **3:**347-350.

14. **Hoshijima, K., K. Inoue, I. Higuchi, H. Sakamoto, and Y. Shimura.** 1991. Control of double-sex alternative splicing by *transformer* and *transformer-2* in Drosophila. Science **252:**833-836.

15. **Kittrell, F. S., C. J. Oborn, and D. Medina.** 1992. Development of mammary preneoplasias in vivo from mouse mammary epithelial cell lines in vitro. Cancer Res. **52::**1924-1932.

16. **Konig, H., H. Ponta, and P. Herrlich.** 1998. Coupling of signal transduction to alternative pre-mRNA splicing by a composite splice regulator. EMBO J. **17:**2904-2913.

17. **Lynch, K. W., and T. Maniatis.** 1996. Assembly of specific SR protein complexes on distinct regulatory elements of the Drosophila doublesex splicing enhancer. Genes Dev. **10:**2089-2101.

18. **Lynch, K. W., and T. Maniatis.** 1995. Synergistic interactions between two distinct elements of a regulated splicing enhancer. Genes Dev. **9:**284-293.

19. **Mackay, C. R., H.-J. Terpe, R. Stauder, W. L. Marston, H. Stark, and U. Gunthert.** 1994. Expression and modulation of CD44 variant isoforms in humans. J. Cell. Biol. **124:**71-82.

20. **Manley, J. L., and R. Tacke.** 1996. SR proteins and splicing control. Genes Dev. **10:**1569-1579.

21. **Medina, D.** 1996. Mammary Tumor Cell Cycle, Differentiation and Metastases, p. 37-69. *In* D. R. a. L. ME (ed.). Kluger Academic Publishers: Norwell.

22. **Peng, X., and S. M. Mount.** 1995. Genetic enhancement of RNA processing defects by a dominant mutation in B52, the Drosophila gene for an SR protein splicing factor. Mol. Cell. Biol. **15:**6273-6282.

23. **Reed, R.** 1996. Initial splice site recognition and pairing during pre-mRNA splicing. Curr. Opin. Genet. Dev. **6:**215-220.

24. **Screaton, G. R., M. V. Bell, D. G. Jackson, F. B. Cornelis, U. Gerth, and J. I. Bell.** 1992. Genomic structure of DNA encoding the lymphocyte homing receptor

CD44 reveals at least 12 alternatively spliced exons. Proc. Natl. Acad. Sci. **89:**12160-12164.

25. **Stamenkovic, I., M. Amiot, J. M. Pesando, and B. Seed.** 1989. A lymphocyte molecule implicated in lymph node homing is a member of the cartilage link protein family. Cell **56:**1057-1062.

26. **Stickeler, E., F. Kittrell, D. Medina, and S. M. Berget.** 1999. Stage-specific changes in SR splicing factors and alternative splicing in mammary tumorigenesis. Oncogene **18:**3574-3582.

27. **Sommerville, J. and M. Ladomery.** 1996. Masking of mRNA by Y-box proteins. FASEB J. **10:**435-443.

28. **Tacke, R., M. Tohyama, S. Ogawa, and J. L. Manley.** 1998. Human Tra 2 proteins are sequence-specific activators of pre-mRNA splicing. Cell **93:**139-148.

29. **Tafuri, S. R., M. Familari, and A. P. Wolffe.** 1993. A mouse Y box protein MSY1 is associated with paternal mRNA in spermatocytes. J. Biol. Chem. **268:**12213-12220.

30. **Tian, M., and T. Maniatis.** 1992. Positive control of pre-mRNA splicing in vitro. Science **256:**237-240.

31. **Wolffe, A. P., S. Tafuri, M. Ranjan, and M. Familari.** 1992. The Y-box factors: A family of nucleic acid binding proteins conserved from *Escherichia coli* to man. The New Biologist **4:**290-298.

32. **Zahler A. M., W. S. Lane, J. A. Stolk, and M. B. Roth.** 1992. SR proteins: a conserved family of pre-mRNA splicing factors. Genes Devel. **6:**837-847.

(11) APPENDICES

Figure Legends

Figure 1. Depiction of the splicing variants of the human CD44 gene. The exon/intron architecture of the human CD44 gene is depicted at top. The ten alternative cassette exons are depicted in red. The exon studied in this report is the fourth variable exon v4. The sequence of this exon is shown at the bottom of the figure. Sequences within the exon are boxed. Red sequence indicate AC-rich elements (ACE 1, ACE 2, ACE 3) within exon v4 that are potential binding sites for trans-acting factors and which have been mutated in this study.

Figure 2. AC-rich repeats within CD44 variable exons 4 and 5. The AC-rich sequences from the two exons are aligned and a consensus repeat sequence (blue) is derived. The three elements within exon 4 are termed ACE 1, ACE 2, and ACE 3 as indicated. At top is the derived repeat consensus from the *D. melanogaster* doublesex gene known to bind to Tra 2 during doublesex alternative splicing..

Figure 3. Human Tra 2α enhances the UV cross-linking of the y-box protein YB1 to CD44 exon 4. (A) Tra 2α stimulates the binding of YB1 to substrate in the absence of other nuclear proteins. Purified recombinant YB1 was incubated with a radiolabeled RNA consisting of the first half of exon 4 (including ACE1 and ACE 2) under splicing conditions but without extract. UV cross-linking was performed after 5 min. of incubation at 30°C in the presence of 2mg/ml heparin. Increasing concentrations of recombinant Tra 2α was added prior to incubation (lanes 2-4). Arrows mark the position

if intact YB1 protein. A second band is visible in lane 1 corresponding to a partial YB1 protein present in the recombinant protein preparation. Cross-linked Tra 2α runs slightly below this band as indicated by the arrow. (B) Tra 2a stimulates the binding of YB1 to exon 4 sequences in the presence of extract proteins. A cross-linking reaction using the substrate in (A) and a splicing S100 extract was performed in the presence of no additional proteins (lane 1), recombinant Tra 2α (lane 2), BSA (lane 3), recombinant polypyrimidine tract binding protein (PTB, lane 4), or a 50-75% ammonium sulfate cut of nuclear extract that contains the major SR proteins SRp75, SRp55, SRp40, ASF/SF2, SC35, and SRp20 (lane 5). The position of YB1 is denoted with an arrow.

Figure 4. Immunoprecipitation of the 50 kDA cross-linked band with antibodies directed against YB1. To formally demonstrate that the 50 kDA protein band observed in UV cross-linking experiments using splicing extract is YB1, proteins from a cross-linking reaction using nuclear extract were immunoprecipitated with an antibody specific for YB1. The displayed gel shows total cross-linked protein (lane 1), the supernatant from the immunoprecipitation (lane 2) and the pellet from the immunoprecipitation (lane 3). Similar results were observed using a reaction with S100 extract (data not shown).

Figure 5. The cross-linking of YB1 to exon is competed by an RNA containing multiple AC-rich sequences known to bind YB1. A competitor RNA for YB1 binding was utilized that contained a sequence known to bind YB1 well (5, diagrammed as the competitor, right). A substrate CD44 RNA was created that contained the first half of exon 4 including ACE 1 and ACE 2. UV Cross-linking to this substrate was performed

in a splicing assay using nuclear extract in the presence of no competitor (lane 1) or increasing concentrations of the AC-rich competitor RNA (lanes 2-4).

Figure 6. Binding of YB1 to wild type and mutant substrates. The binding of YB1 to three substrates was compared by UV cross-linking in the presence of competitor RNA and immunoprecipitation with antibodies specific for YB1. Substrates include wild type exon 4 containing ACE1 and ACE 2 as depicted in Figure 5 (lanes 2-6), A similar substrate in which the ACE 1 had been mutated (lanes 7-11; mutation drawn in Figure 11), and the AC-rich sequence described in Figure 5 which is also the competitor used in indicated lanes. The position of YB1 is indicated.

Figure 7. YB1 and Tra 2 can bind exon 4 simultaneously. The binding of YB1, Tra 2α and Tra 2β was analyzed by gel shift analysis. Radiolabeled substrate containing ACE 1 and ACE 2 was incubated with the indicated recombinant proteins under splicing conditions but without extract and with 2 mg/ml heparin for 10 min. at 30°C. Formed complexes were analyzed on a neutral 5% acrylamide gel. The migration position of free RNA is indicated at the left. The complexes formed with substrate and YB1, YB1 + Tra 2α, or YB1, + Tra 2β are indicated with arrows.

Figure 8. Increasing the in vivo concentration of Tra 2α increased the inclusion of CD44 variable exons 4 and 5. The diagrammed mini-gene in which CD44 variable exons 4 and 5 along with their flanking introns were inserted into the human beta-globin gene was transfected into HeLa cells with or without an expression plasmid coding for

human Tra 2α. Splicing phenotypes were determined by RT/PCR using primers specific for the flanking beta-globin sequences. (A) Quantification of multiple experiments. Results from RT/PCR amplifications using radiolabeled primers were compared by quantification on the Phosphoimager. Results are plotted as the percentage of mRNA including exons v4 and v5. One standard deviation is shown. (B) Diagram of the mini-gene employed. Black indicates beta-globin sequences, red indicates CD44 sequences. (C) Representative experiment. RT/PCR amplification of RNA produced after transfection of the indicated mini-gene and expression plasmids coding for human Tra 2α (left) or a control SR protein that does not affect CD44 alternative splicing (right). Increasing amounts of expression plasmids were used as indicated varying from 1 to 4 μg of plasmid per 10 cm dish of cells.

Figure 9. Transfection of human YB1 increased the inclusion of CD44 variable exons 4 and 5. An experiment similar to that in Figure 7 was performed in which the co-transfecting expression plasmids coded for either wild type YB1 or a truncation mutant of YB1 that ended in the middle of the charged protein-protein interaction domain of the protein. Increasing amounts (0-4 mg) of the expression plasmid were used as indicated. The quantification is from multiple experiments. The cartoon at bottom right shows the structure of the utilized wild type and YB1 genes.

Figure 10. YB1 does not affect splicing of a heterologous weak exon in the beta-globin mini-gene backbone or expression levels of a cDNA containing v4 and v5. (A) Expression of a cDNA including v4 and v5 in the presence and absence of a YB1

expression plasmid. A cDNA construct was prepared that was identical to that used in Figures 8 and 9 except that the introns were removed (i.e. identical to the spliced product including v4 and v5). This construct was transfected into HeLa cells with or without the expression vector coding for YB1 and resulting RNA was quantified by RT/PCR using the same primers employed in Figures 8 and 9. An arrow marks the position of intact cDNA. (B) Effect of YB1 and Tra 2α on splicing of a weak beta globin pre-mRNA. The diagrammed beta-globin gene was transfected into HeLa cells with or without expression plasmids coding for YB1 or Tra 2α. The mini-gene is a version of beta-globin containing an internally shortened middle exon which is known to be weak and included only 30% of the time. This exon has been deleted for all known globin enhancer sequences and represents a weak constitutive exon without enhancing sequences.

Figure 11. Mutation of ACE 1 and ACE 3 alters the *in vivo* splicing exon v4. Mutant versions of the *in vivo* expression mini-gene used in Figures 8 and 9 were constructed containing the indicated mutations in exon v4 ACE 1, ACE 2, or ACE 3. Mutated nucleotides are indicated in red. Increasing amounts of these DNAs were transfected into HeLa cells and the splicing pattern was determined by RT/PCR. The position of RNA containing neither CD44 exon, either exon v4 or v5 but not both, or both v4 and v5 are shown. Note that because exons v4 and v5 are almost the same size it is impossible to determine by RT/PCR which exon is represented by the middle amplification product. Sequencing of the band produced with the ACE 3 mutant indicates that this RNA product contains exon v5 but not v4.

Figure 12. Mutation of ACE 3 inhibits *in vitro* assembly of ATP-dependent spliceosome complex A on substrates consisting of exon v4. A substrate RNA consisting of the complete exon v4 and its flanking splice sites was incubated in a standard splicing assay. Spliceosome assemblies were displayed on neutral RNP gels. The position of the complexes resulting from the binding of hnRNP proteins (complex H) or the binding of U1 and U2 snRNPs plus other splicing factors (complex A) are indicated. Mutant substrates containing the mutations diagrammed in Figure 11 were similar analyzed. The assembly pathway for substrates of this exon/intron architecture is indicated to the left.

Figure 13. Mutation of ACE 1 inhibits UV cross-linking of YB1 to exon v4. Radiolabeled substrate RNAs consisting of the first half of exon v4 and either wild type or mutant ACE 1 and ACE 2 as indicated were assayed in a standard in vitro UV cross-linking assay using nuclear extract (lanes 1-6) or S100 extract (lanes 7-11). Also shown is the cross-linking of an RNA containing the AC-rich YB1 binding sequence described in Figures 5 and 6. The position of YB1 is indicated.

Figure 14. YB1 is induced during murine mammary development and tumorigenesis. RNA was isolated from normal murine mammary gland (lanes 1-4, 7), pregnant gland (lane 5), lactating gland (lane 6), independent mammary gland preneoplasias (lanes 8 and 9), independent tumors (lanes 10-11), and a lung metastases of a mammary tumor (lane 12) (23). RNA from transfected cells was analyzed by RT/PCR amplification using primers specific for murine YB1 (29). Products were displayed on

agarose gels and stained with ethidium bromide. RNA preparations were checked for intact RNA by RT/PCR analysis of control RNAs.

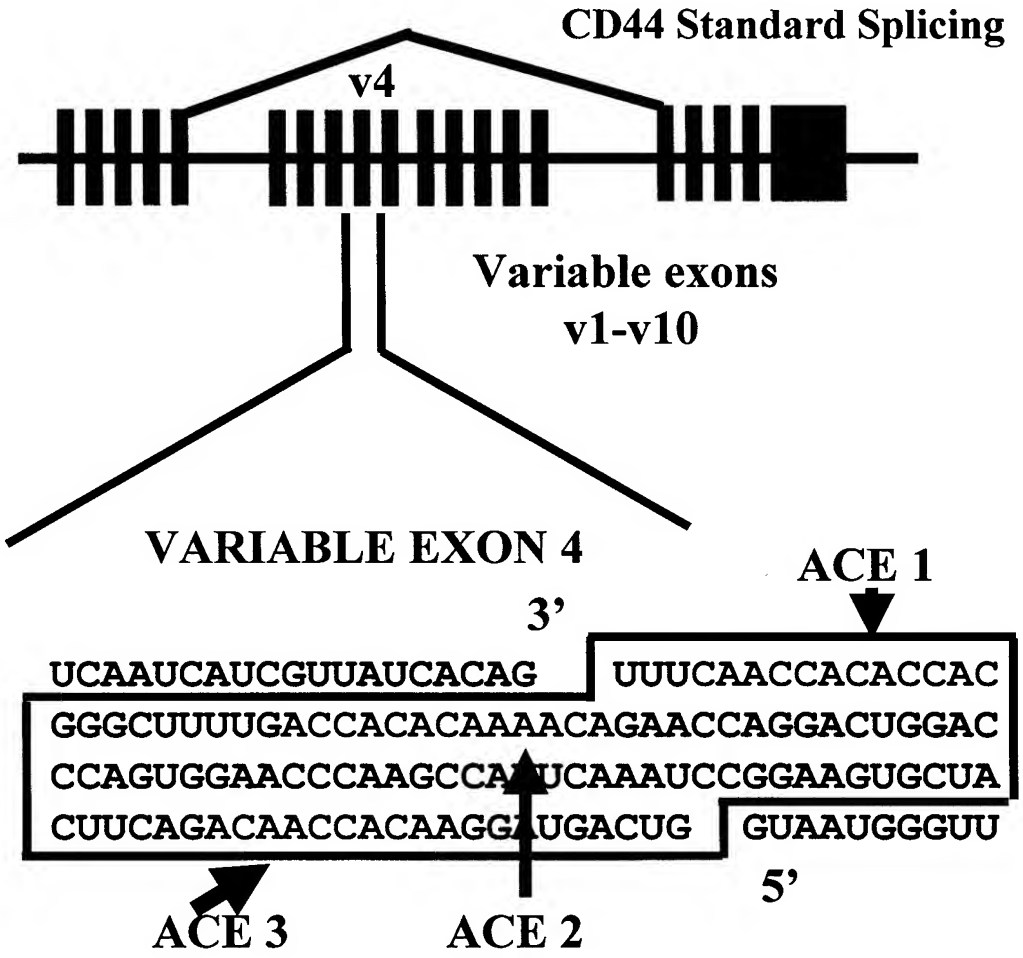

CD44 Standard Splicing
v4
Variable exons
v1-v10
VARIABLE EXON 4
ACE 1
3'
UCAAUCAUCGUUAUCACAG
UUUCAACCACACCAC
GGGCUUUUGACCACACAAAACAGAACCAGGACUGGAC
CCAGUGGAACCCAAGCCA
UCAAAUCCGGAAGUGCUA
CUUCAGACAACCACAAGG
UGACUG
GUAAUGGGUU
5'
ACE 3
ACE 2

Figure 1

CD44 exons v4 & v5 have AC-rich sequences similar to those known to bind Tra 2

AC-rich Repeats in CD44 Variable Exons 4 & 5

CAAUCAACAUU	D. melanogaster doublesex repeat
CAAUCAUCGUU	CD44 exon 4 3' splice site
UAACCAUCAUC	CD44 exon 5 3' splice site
CAACCA CACCA	CD44 exon 4 sequence 1 ACE 1
GACCA CACAA	CD44 exon 4 sequence 2 ACE 2
CAACCA CAA	CD44 exon 4 sequence 3 ACE 3
CAAGCA CAA	CD44 exon 5 sequence 1
CAACCA CA	CD44 Consensus

Figure 2

Human Tra 2α enhances UV cross-linking of Y-box Protein YB1 to CD44 exon 4

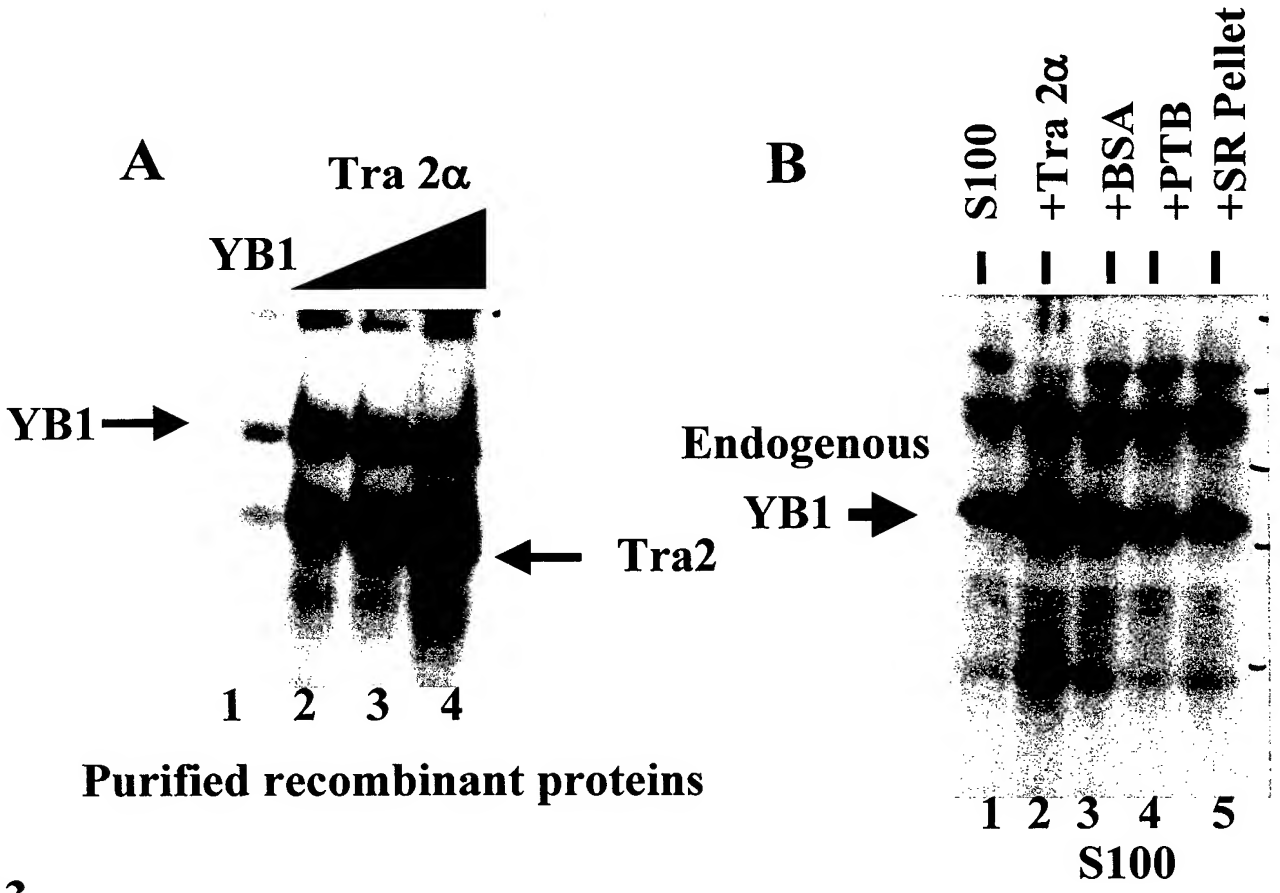


Figure 3

CD44 exon 4 binds YB1

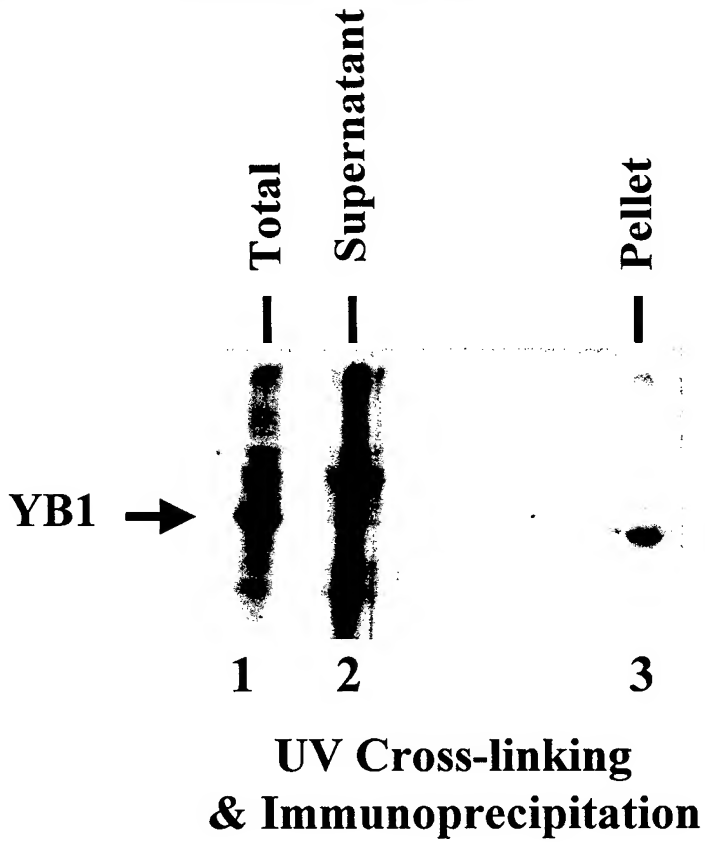
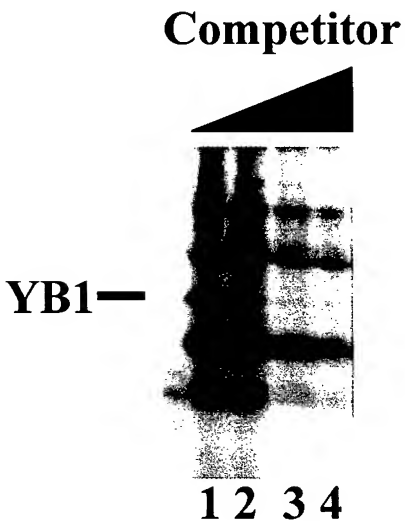


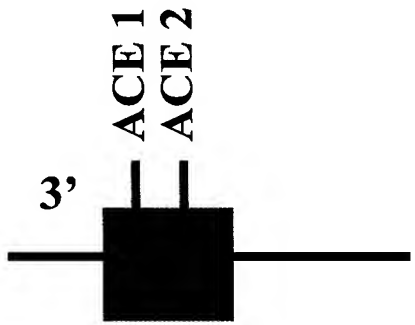
Figure 4



Competitor:
CACCA....CCACCA

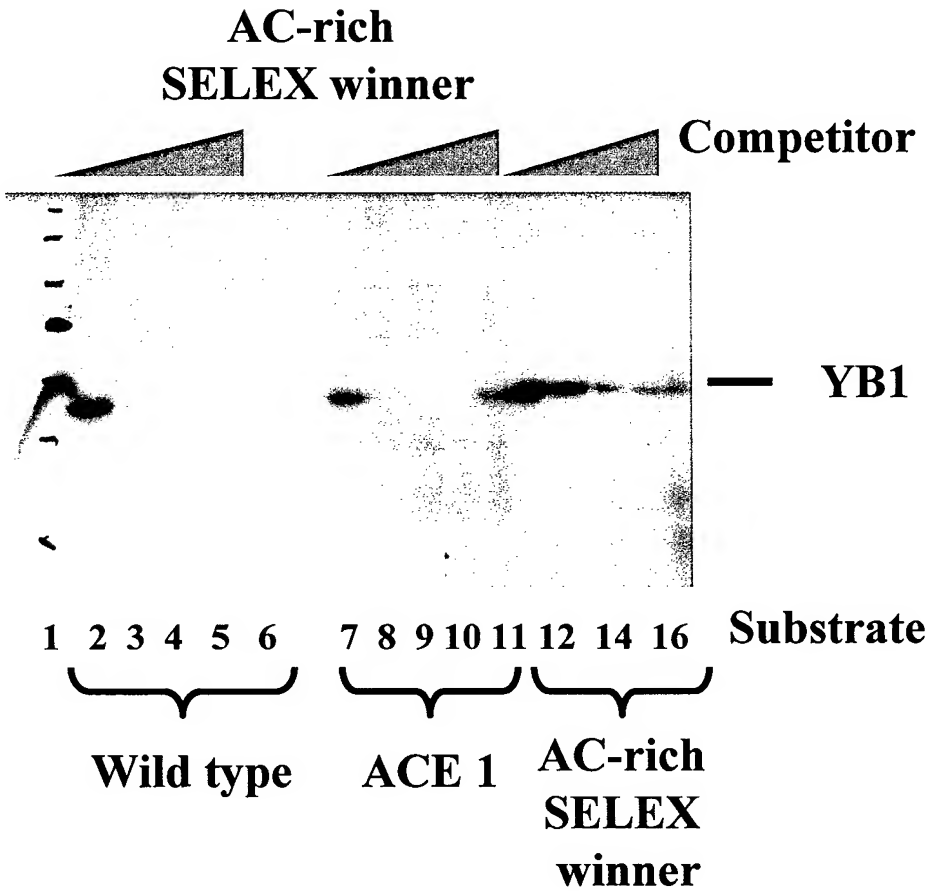
ACE 1 CAACCACACCAC

ACE 2 ACCACACAAAACA



Exon 4
(nt 1 - 49)

Figure 5



The binding of both wild type and ACE 1 RNAs to YB1 is easily competed with an RNA known to bind YB1

Figure 6

Gel-shift analysis indicates the presence of both YB1 and Tra 2 in complexes with CD44 exon 4

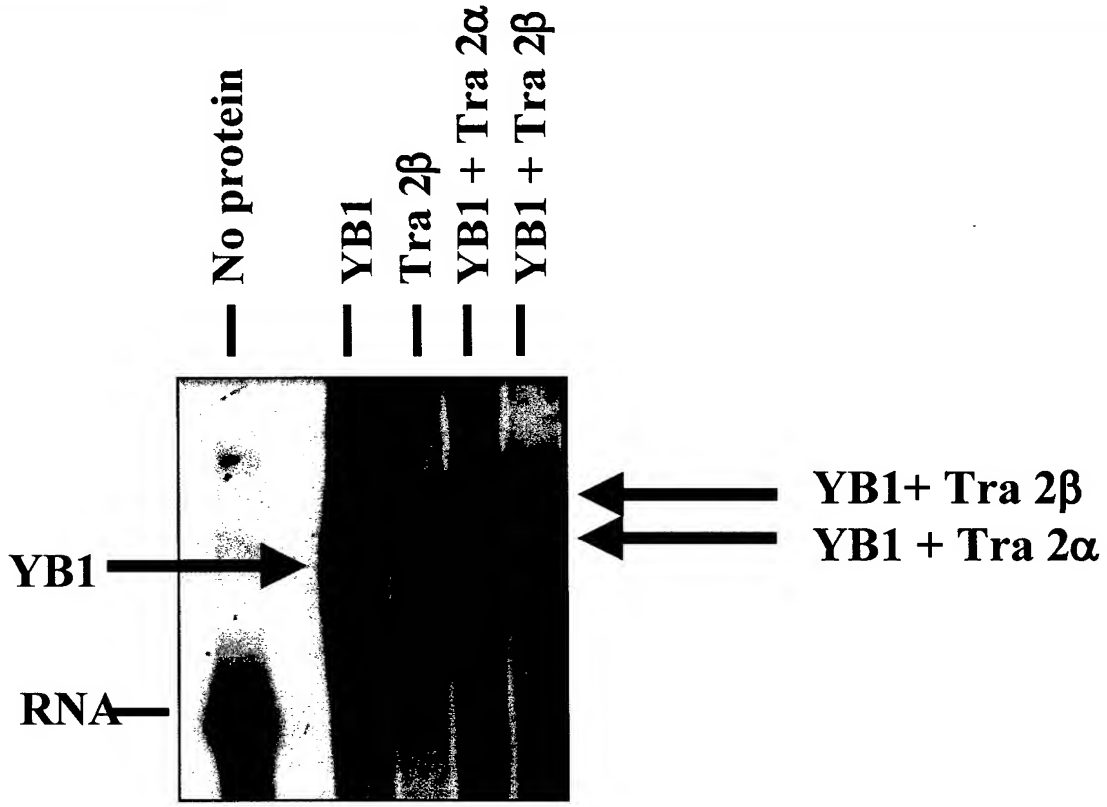


Figure 7

Human Tra 2α Enhanced Inclusion of CD44 Variable Exons 4 & 5

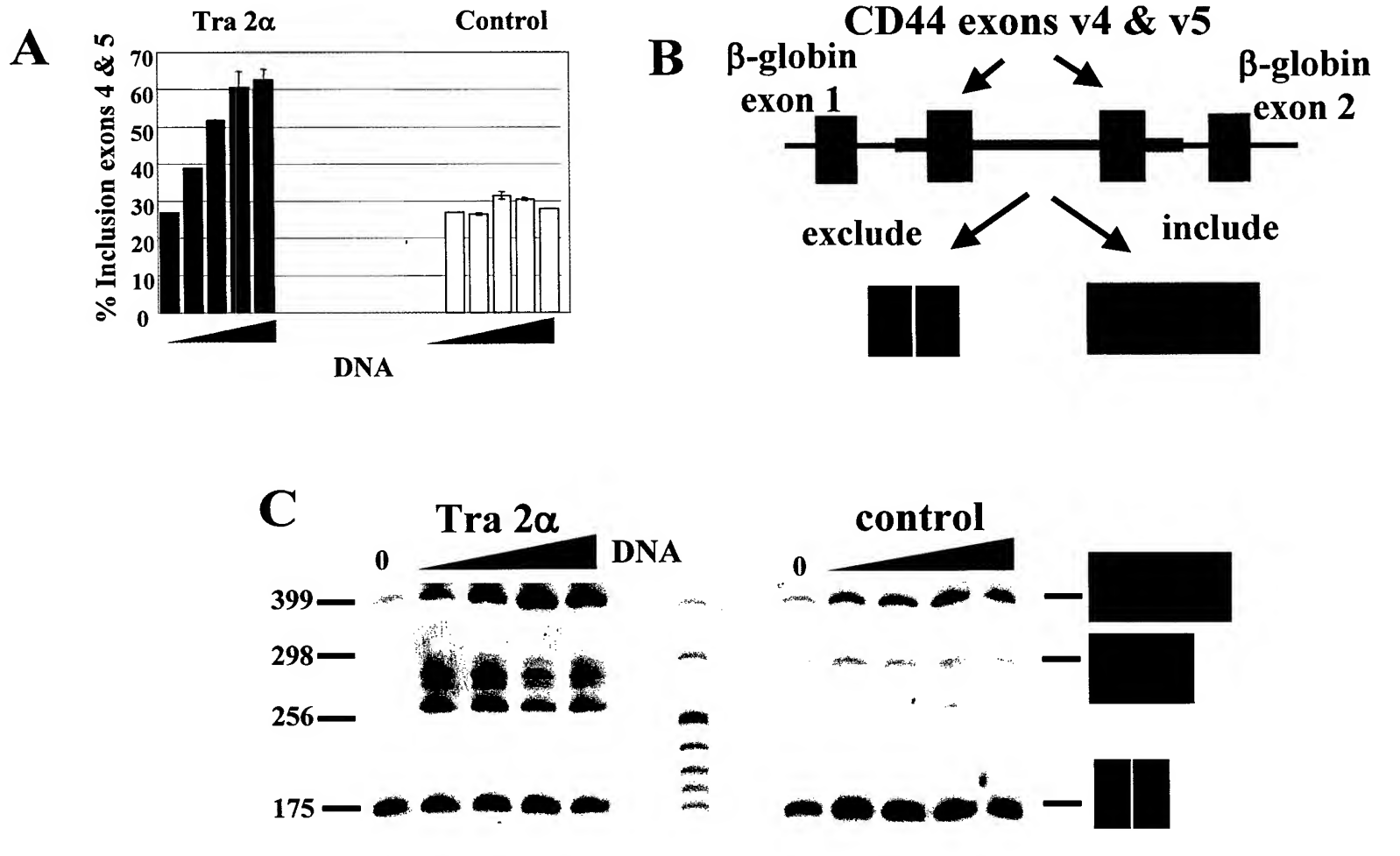
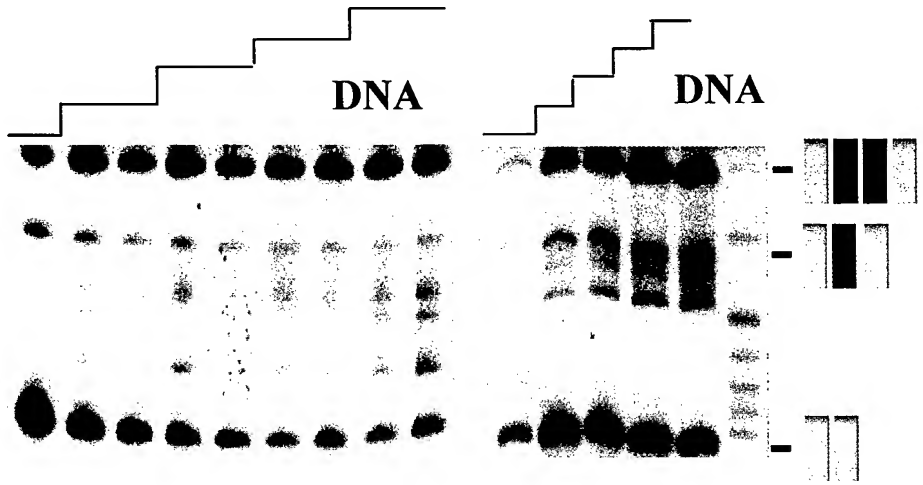
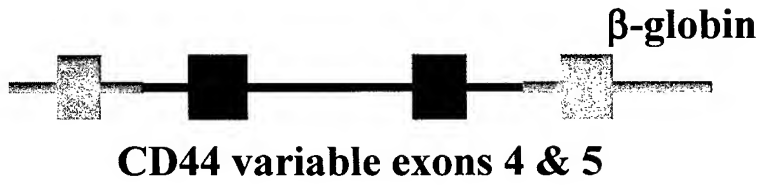
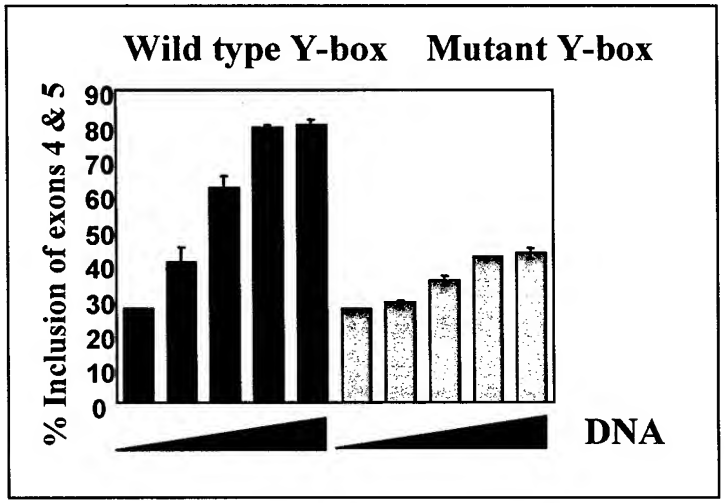


Figure 8

Co-transfection with YB1 increased inclusion of CD44 variable exons 4 & 5







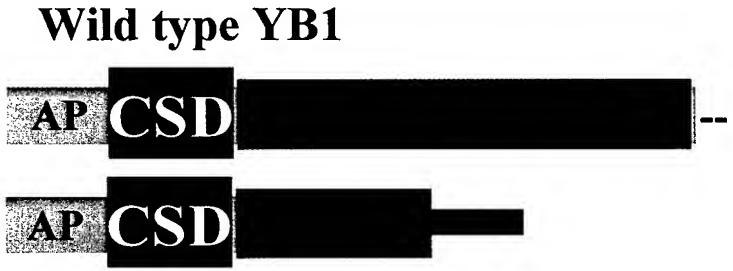
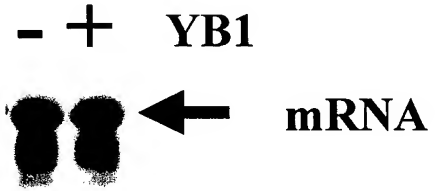
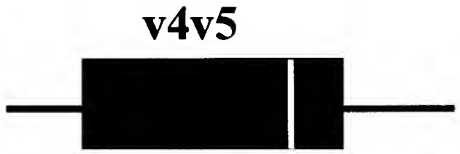


Figure 9

A

YB1 does not affect expression of a cDNA containing v4 and v5

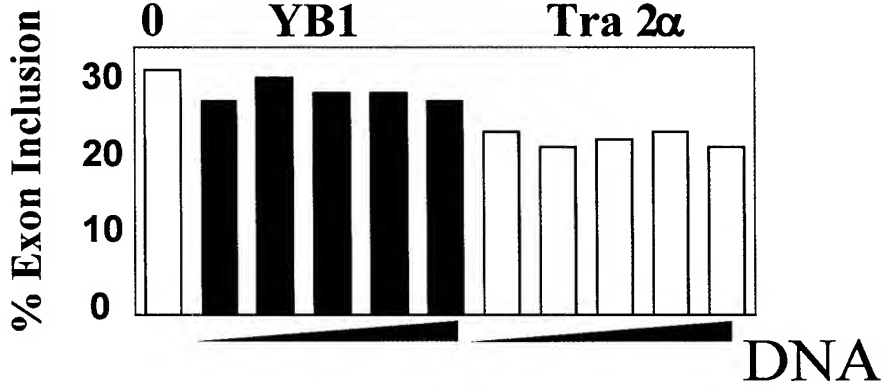


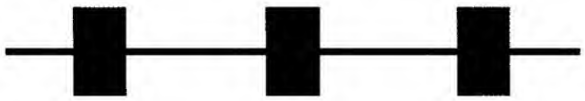


CD44 mini-gene cDNA

B

Neither YB1or human Tra 2α affect inclusion of a weak globin exon

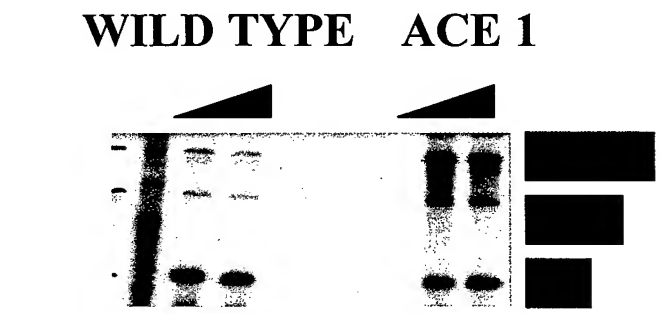




β-globin dup33

Figure 10

Mutation of ACE 1 and ACE 3 but not ACE 2 lowers *in vivo* inclusion of v4 or v4 and v5



ACE 1 CAACCACACCAC
GG GG

ACE 2 ACCACACAAAACA
G G T

ACE 3 CAACCACAA
T GG T

Figure 11

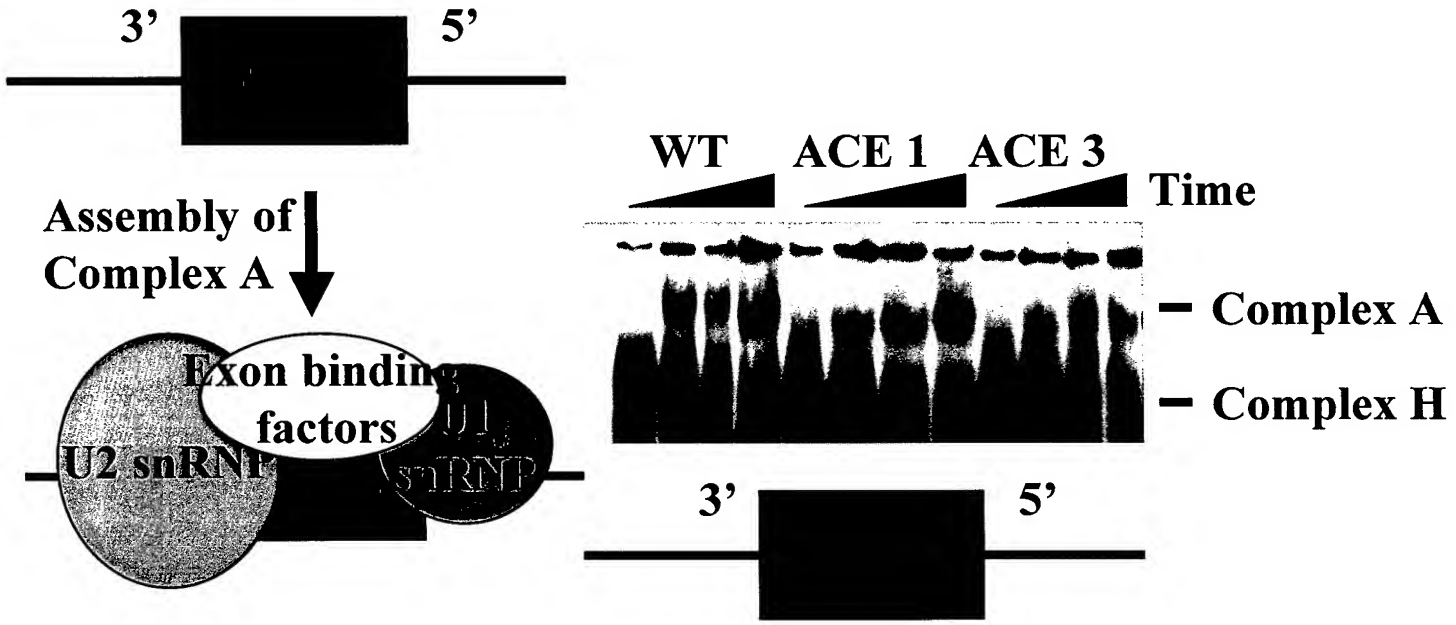
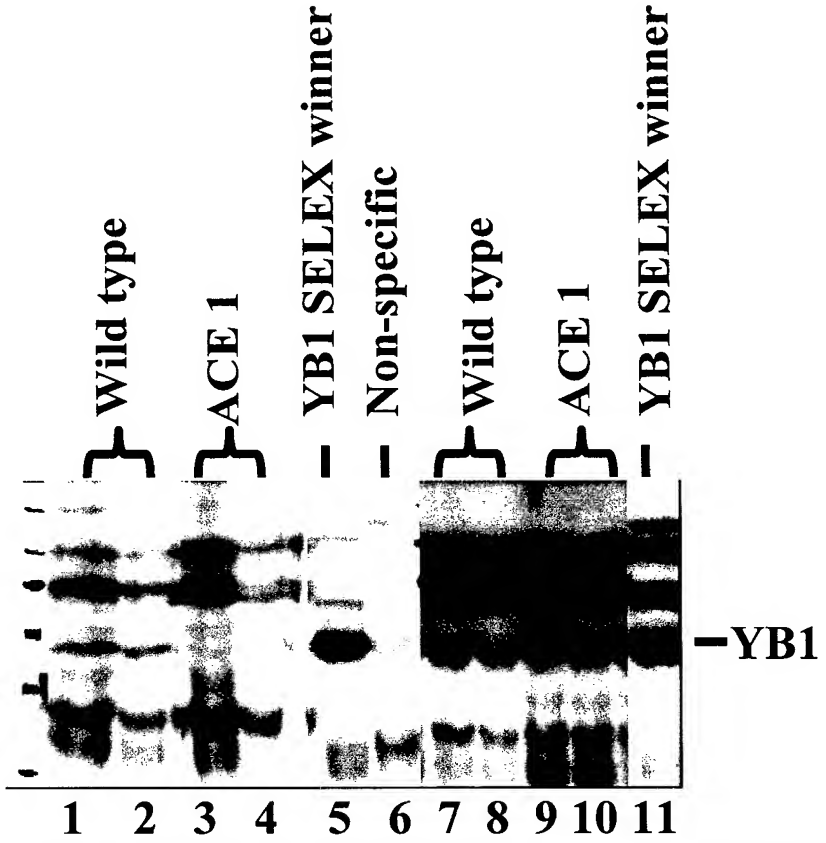

3'
5'
Assembly of
Complex A
Exon binding
factors
U2 snRNP
U1
snRNP
WT
ACE 1
ACE 3
Time
Complex A
Complex H
3'
5'

Figure 12



The ACE 1 Phenotype disappears in Extract made from a cytoplasmic S100

Figure 13

YB1 is Induced in Mammary Development and Tumorigenesis

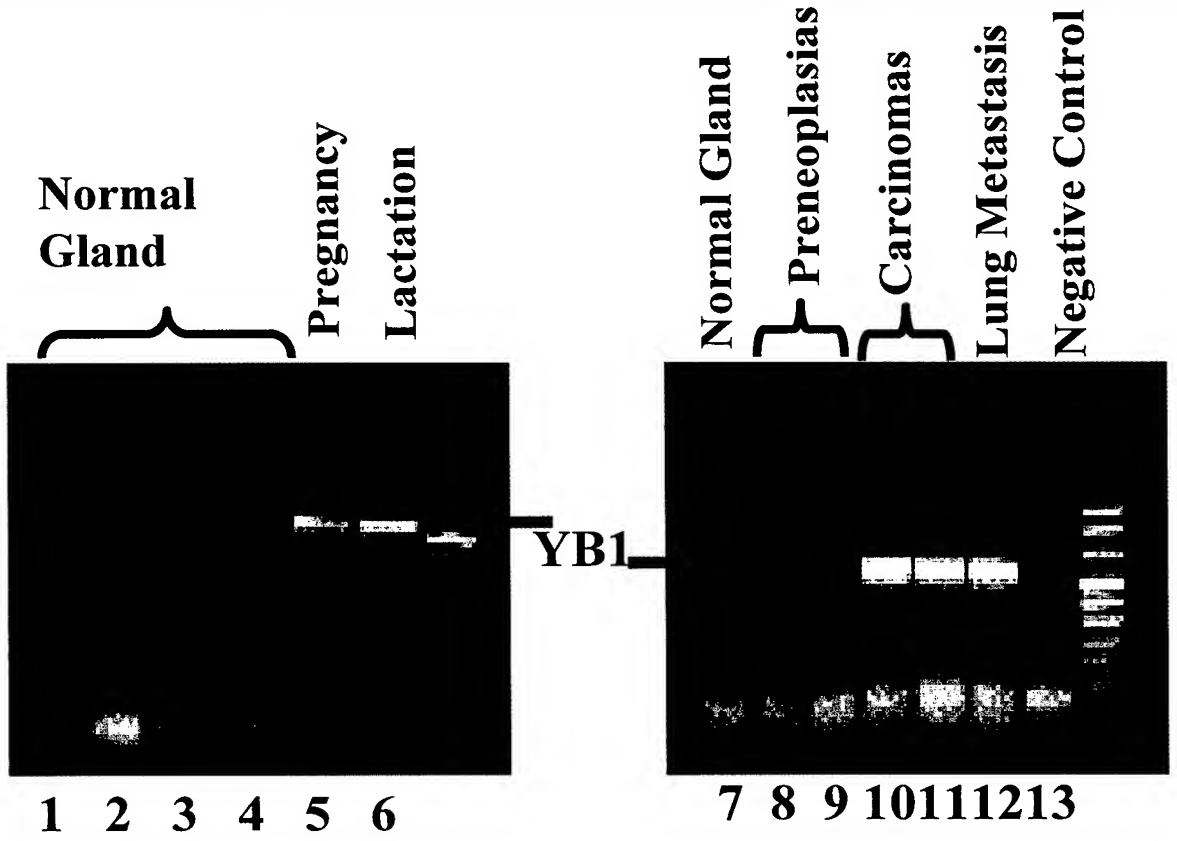


Figure 14